

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7590472号
(P7590472)

(45)発行日 令和6年11月26日(2024.11.26)

(24)登録日 令和6年11月18日(2024.11.18)

(51)国際特許分類

F I

G 0 6 Q 30/0201(2023.01)

G 0 6 Q 30/0201

請求項の数 13 (全27頁)

(21)出願番号	特願2023-11858(P2023-11858)	(73)特許権者	507009009
(22)出願日	令和5年1月30日(2023.1.30)		株式会社博報堂D Yホールディングス
(65)公開番号	特開2024-107764(P2024-107764		東京都港区赤坂五丁目3番1号
	A)	(74)代理人	110000578
(43)公開日	令和6年8月9日(2024.8.9)		名古屋国際弁理士法人
審査請求日	令和5年9月7日(2023.9.7)	(72)発明者	大杉 巧
			東京都港区赤坂五丁目3番1号 株式会
			社博報堂D Yホールディングス内
		(72)発明者	熊谷 雄介
			東京都港区赤坂五丁目3番1号 株式会
			社博報堂D Yホールディングス内
		(72)発明者	木田 裕一郎
			東京都港区赤坂五丁目3番1号 株式会
			社博報堂D Yホールディングス内
		(72)発明者	道本 龍
			最終頁に続く

(54)【発明の名称】 情報処理システム、コンピュータプログラム、及び情報処理方法

(57)【特許請求の範囲】

【請求項1】

第一の標本としての複数の第一の対象者に関する第一のデータセットを取得する第一の取得部と、

第二の標本としての複数の第二の対象者に関する第二のデータセットを取得する第二の取得部と、

前記第二のデータセットに基づき、前記第一のデータセットを補正する補正部と、
を備え、

前記第一のデータセットは、前記複数の第一の対象者のそれぞれに関して、対応する第一の対象者の行動を説明する行動データを、第一の行動データとして備え、

前記第二のデータセットは、前記複数の第二の対象者のそれぞれに関して、対応する第二の対象者の行動を説明する行動データを、第二の行動データとして備え、

前記第一の行動データは、複数の第一の行動に関して、第一の行動毎に、前記対応する第一の対象者の、対応する行動を数値により説明し、

前記第二の行動データは、複数の第二の行動に関して、第二の行動毎に、前記対応する第二の対象者の、対応する行動を数値により説明し、

前記複数の第二の行動には、前記複数の第一の行動が含まれ、

前記第一の標本は、特定グループに属する人に限定された人の集合であり、

前記第二の標本は、前記特定グループに属さない人の集合又は前記特定グループに属する人に限定されない人の集合であり、

前記第一の行動データと前記第二の行動データとの間では、前記行動の説明のために共通する変数である共通変数が用いられ、

前記補正部は、

前記第一のデータセット及び前記第二のデータセットに基づく前記共通変数で条件付けられる人が前記特定グループに所属する確率に対応する傾向スコアに基づき、前記第一の標本が前記特定グループに属する人の集合に限られていることに起因する前記第一のデータセットにおける前記数値のバイアスである第一のバイアスを、前記第一の対象者毎に特定し、

前記第一のデータセットにおける前記複数の第一の対象者間の行動環境の相違に起因する前記数値のバイアスである第二のバイアスを、前記第一の対象者及び前記第一の行動の組み合わせ毎に特定し、

前記第一のデータセットにおける前記複数の第一の対象者に関する前記第一の行動毎の前記数値を、前記第一のバイアスの逆数及び前記第二のバイアスの逆数で補正することにより、前記第一のバイアス及び前記第二のバイアスを抑制する方向に、前記第一のデータセットを補正し、

前記第一のバイアスは、前記第一の対象者の前記共通変数での前記傾向スコアに基づく前記第一の対象者の前記第一の行動が観測される可能性の大きさに対応し、

前記第二のバイアスは、前記組み合わせに対応する第一の対象者が置かれている環境において前記組み合わせに対応する第一の行動が観測される因子の大きさに対応する情報処理システム。

【請求項 2】

第一の標本としての複数の第一の対象者に関する第一のデータセットを取得する第一の取得部と、

第二の標本としての複数の第二の対象者に関する第二のデータセットを取得する第二の取得部と、

前記第二のデータセットに基づき、前記第一のデータセットを補正する補正部と、
を備え、

前記第一のデータセットは、前記複数の第一の対象者のそれぞれに関して、対応する第一の対象者の行動を説明する行動データを、第一の行動データとして備え、

前記第二のデータセットは、前記複数の第二の対象者のそれぞれに関して、対応する第二の対象者の行動を説明する行動データを、第二の行動データとして備え、

前記第一の行動データは、複数の第一の行動に関して、第一の行動毎に、前記対応する第一の対象者の、対応する行動を数値により説明し、

前記第二の行動データは、複数の第二の行動に関して、第二の行動毎に、前記対応する第二の対象者の、対応する行動を数値により説明し、

前記複数の第二の行動には、前記複数の第一の行動が含まれ、

前記第一の標本は、特定グループに属する人に限定された人の集合であり、

前記第二の標本は、前記特定グループに属さない人の集合又は前記特定グループに属する人に限定されない人の集合であり、

前記第一の行動データ及び前記第二の行動データは、前記行動の主体である対象者の属性を、少なくとも一つの変数で表す属性情報を有し、

前記第一の行動データと前記第二の行動データとの間では、前記属性の説明のために共通する変数である共通変数が用いられ、

前記補正部は、

前記第一のデータセット及び前記第二のデータセットに基づく前記共通変数で条件付けられる人が前記特定グループに所属する確率に対応する傾向スコアに基づき、前記第一の標本が前記特定グループに属する人の集合に限られていることに起因する前記第一のデータセットにおける前記数値のバイアスである第一のバイアスを、前記第一の対象者毎に特定し、

前記第一のデータセットにおける前記複数の第一の対象者間の行動環境の相違に起因す

10

20

30

40

50

る前記数値のバイアスである第二のバイアスを、前記第一の対象者及び前記第一の行動の組み合わせ毎に特定し、

前記第一のデータセットにおける前記複数の第一の対象者に関する前記第一の行動毎の前記数値を、前記第一のバイアスの逆数及び前記第二のバイアスの逆数で補正することにより、前記第一のバイアス及び前記第二のバイアスを抑制する方向に、前記第一のデータセットを補正し、

前記第一のバイアスは、前記第一の対象者の前記共通変数での前記傾向スコアに基づく前記第一の対象者の前記第一の行動が観測される可能性の大きさに対応し、

前記第二のバイアスは、前記組み合わせに対応する第一の対象者が置かれている環境において前記組み合わせに対応する第一の行動が観測される因子の大きさに対応する情報処理システム。

10

【請求項 3】

前記複数の第一の行動及び前記複数の第二の行動のそれぞれは、購買行動であり、

前記第一の標本は、購買に関する特定のリワードプログラムに参加する消費者、又は、特定の電子決済サービスを利用する消費者の集合であり、

前記第一の行動データは、前記対応する第一の対象者からの識別情報の提供を通じて観測された前記対応する第一の対象者の購買行動を説明するデータである請求項 1 又は請求項 2 記載の情報処理システム。

【請求項 4】

前記複数の第一の行動及び前記複数の第二の行動のそれぞれは、購買対象及び購買店舗の少なくとも一方によって区別される購買行動である請求項 1 又は請求項 2 記載の情報処理システム。

20

【請求項 5】

前記複数の第一の行動及び前記複数の第二の行動のそれぞれは、情報端末を通じた行動である請求項 1 又は請求項 2 記載の情報処理システム。

【請求項 6】

前記補正部は、複数の地域に関して、地域毎に、前記複数の第一の行動のそれぞれの実行機会の大小を判別し、前記第一の対象者及び前記第一の行動の組み合わせ毎に、前記組み合わせに対応する第一の対象者の活動地域における、前記組み合わせに対応する第一の行動の前記実行機会の大小に起因する前記第二のバイアスを特定し、

30

前記第二のバイアスは、前記実行機会の大きさに対応する請求項 1 又は請求項 2 記載の情報処理システム。

【請求項 7】

前記複数の第一の行動及び前記複数の第二の行動のそれぞれは、少なくとも購買店舗によって区別される購買行動であって、複数の店舗のうちの一つの店舗における購買行動であり、

前記一つの店舗には、同系の店舗群が含まれ、

前記補正部は、複数の地域に関して、地域毎に、前記複数の店舗のそれぞれのシェアを判別し、前記第一の対象者及び前記第一の行動の組み合わせ毎に、前記組み合わせに対応する第一の対象者の活動地域における、前記組み合わせに対応する第一の行動が行われた店舗のシェアに起因する前記第二のバイアスを特定し、

40

前記第二のバイアスは、前記シェアの大きさに対応する請求項 1 又は請求項 2 記載の情報処理システム。

【請求項 8】

前記複数の第一の行動及び前記複数の第二の行動のそれぞれは、少なくとも購買店舗によって区別される購買行動であって、複数の店舗のうちの一つの店舗における購買行動であり、

前記補正部は、前記複数の店舗に関して、店舗毎に、対応する店舗における特定決済手段の消費者属性毎の利用量を判別し、前記第一の対象者及び前記第一の行動の組み合わせ毎に、前記組み合わせに対応する第一の対象者の消費者属性に関する、前記組み合わせに

50

対応する第一の行動が行われた店舗における特定決済手段の利用量に起因する前記第二のバイアスを特定し、

前記第二のバイアスは、前記利用量の大きさに対応する請求項 1 又は請求項 2 記載の情報処理システム。

【請求項 9】

前記複数の第一の行動及び前記複数の第二の行動のそれぞれは、情報端末において利用されるアプリケーションソフトウェアの種類によって区別される行動であって、複数種類のアプリケーションソフトウェアのうちの一種類のアプリケーションソフトウェアを用いた行動であり、

前記補正部は、情報端末の種類毎に、対応する種類の情報端末におけるアプリケーションソフトウェアの種類毎のシェアを判別し、前記第一の対象者及び前記第一の行動の組み合わせ毎に、前記組み合わせに対応する第一の対象者が利用する種類の情報端末における、前記組み合わせに対応する前記第一の行動において利用されるアプリケーションソフトウェアのシェアに起因する前記第二のバイアスを特定し、

前記第二のバイアスは、前記シェアの大きさに対応する請求項 1 又は請求項 2 記載の情報処理システム。

【請求項 10】

前記複数の第二の行動には、前記複数の第一の行動とは異なる複数の第三の行動が含まれ、

前記第二の行動データは、前記第一の行動データにおいて説明されない前記複数の第三の行動に関して、前記第三の行動毎に、前記対応する第二の対象者の、対応する行動を数値により説明し、

前記補正部は、

前記第二のデータセットに基づき、前記複数の第一の行動に関する前記数値から前記複数の第三の行動に関する前記数値の推定値を算出するための推定モデルを構築し、

前記複数の第一の対象者のそれぞれに関して、前記推定モデルに基づいて算出した、前記対応する第一の対象者の前記複数の第三の行動に関する前記推定値を、前記第一の行動データに追加することにより、前記第一のデータセットに欠損する前記複数の第三の行動に関する情報を補完するように、前記第一のデータセットを補正し、

前記推定モデルに基づく前記推定値の算出は、前記第一のバイアス及び前記第二のバイアスを抑制する方向に補正された、前記第一の行動データにおける前記複数の第一の行動に関する前記数値に基づいて行われる請求項 1 又は請求項 2 記載の情報処理システム。

【請求項 11】

請求項 1 又は請求項 2 記載の情報処理システムにおける前記第一の取得部と、前記第二の取得部と、前記補正部としての機能を、コンピュータに実現させるためのコンピュータプログラム。

【請求項 12】

コンピュータにより実行される情報処理方法であって、

第一の標本としての複数の第一の対象者に関する第一のデータセットを取得することと、
第二の標本としての複数の第二の対象者に関する第二のデータセットを取得することと、
前記第二のデータセットに基づき、前記第一のデータセットを補正することと、
を含み、

前記第一のデータセットは、前記複数の第一の対象者のそれぞれに関して、対応する第一の対象者の行動を説明する行動データを、第一の行動データとして備え、

前記第二のデータセットは、前記複数の第二の対象者のそれぞれに関して、対応する第二の対象者の行動を説明する行動データを、第二の行動データとして備え、

前記第一の行動データは、複数の第一の行動に関して、第一の行動毎に、前記対応する第一の対象者の、対応する行動を数値により説明し、

前記第二の行動データは、複数の第二の行動に関して、第二の行動毎に、前記対応する第二の対象者の、対応する行動を数値により説明し、

10

20

30

40

50

前記複数の第二の行動には、前記複数の第一の行動が含まれ、
前記第一の標本は、特定グループに属する人に限定された人の集合であり、
前記第二の標本は、前記特定グループに属さない人の集合又は前記特定グループに属する人に限定されない人の集合であり、
前記第一の行動データと前記第二の行動データとの間では、前記行動の説明のために共通する変数である共通変数が用いられ、
 前記補正することは、
前記第一のデータセット及び前記第二のデータセットに基づく前記共通変数で条件付けられる人が前記特定グループに所属する確率に対応する傾向スコアに基づき、前記第一の標本が前記特定グループに属する人の集合に限られていることに起因する前記第一のデータセットにおける前記数値のバイアスである第一のバイアスを、前記第一の対象者毎に特定し、
前記第一のデータセットにおける前記複数の第一の対象者間の行動環境の相違に起因する前記数値のバイアスである第二のバイアスを、前記第一の対象者及び前記第一の行動の組み合わせ毎に特定し、
前記第一のデータセットにおける前記複数の第一の対象者に関する前記第一の行動毎の前記数値を、前記第一のバイアスの逆数及び前記第二のバイアスの逆数で補正することにより、前記第一のバイアス及び前記第二のバイアスを抑制する方向に、前記第一のデータセットを補正することを含み、
前記第一のバイアスは、前記第一の対象者の前記共通変数での前記傾向スコアに基づく前記第一の対象者の前記第一の行動が観測される可能性の大きさに対応し、
前記第二のバイアスは、前記組み合わせに対応する第一の対象者が置かれている環境において前記組み合わせに対応する第一の行動が観測される因子の大きさに対応する情報処理方法。

【請求項 13】

コンピュータにより実行される情報処理方法であって、
 第一の標本としての複数の第一の対象者に関する第一のデータセットを取得することと、
 第二の標本としての複数の第二の対象者に関する第二のデータセットを取得することと、
 前記第二のデータセットに基づき、前記第一のデータセットを補正することと、
 を含み、
 前記第一のデータセットは、前記複数の第一の対象者のそれぞれに関して、対応する第一の対象者の行動を説明する行動データを、第一の行動データとして備え、
 前記第二のデータセットは、前記複数の第二の対象者のそれぞれに関して、対応する第二の対象者の行動を説明する行動データを、第二の行動データとして備え、
 前記第一の行動データは、複数の第一の行動に関して、第一の行動毎に、前記対応する第一の対象者の、対応する行動を数値により説明し、
 前記第二の行動データは、複数の第二の行動に関して、第二の行動毎に、前記対応する第二の対象者の、対応する行動を数値により説明し、
 前記複数の第二の行動には、前記複数の第一の行動が含まれ、
前記第一の標本は、特定グループに属する人に限定された人の集合であり、
前記第二の標本は、前記特定グループに属さない人の集合又は前記特定グループに属する人に限定されない人の集合であり、
前記第一の行動データ及び前記第二の行動データは、前記行動の主体である対象者の属性を、少なくとも一つの変数で表す属性情報を有し、
前記第一の行動データと前記第二の行動データとの間では、前記属性の説明のために共通する変数である共通変数が用いられ、
 前記補正することは、
前記第一のデータセット及び前記第二のデータセットに基づく前記共通変数で条件付けられる人が前記特定グループに所属する確率に対応する傾向スコアに基づき、前記第一の標本が前記特定グループに属する人の集合に限られていることに起因する前記第一のデー

10

20

30

40

50

タセットにおける前記数値のバイアスである第一のバイアスを、前記第一の対象者毎に特定し、

前記第一のデータセットにおける前記複数の第一の対象者間の行動環境の相違に起因する前記数値のバイアスである第二のバイアスを、前記第一の対象者及び前記第一の行動の組み合わせ毎に特定し、

前記第一のデータセットにおける前記複数の第一の対象者に関する前記第一の行動毎の前記数値を、前記第一のバイアスの逆数及び前記第二のバイアスの逆数で補正することにより、前記第一のバイアス及び前記第二のバイアスを抑制する方向に、前記第一のデータセットを補正することを含み、

前記第一のバイアスは、前記第一の対象者の前記共通変数での前記傾向スコアに基づく前記第一の対象者の前記第一の行動が観測される可能性の大きさに対応し、

前記第二のバイアスは、前記組み合わせに対応する第一の対象者が置かれている環境において前記組み合わせに対応する第一の行動が観測される因子の大きさに対応する情報処理方法。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、情報処理システム及び情報処理方法に関する。

【背景技術】

【0002】

人の行動を分析することが従来行われている。例えば、消費者の購買行動を分析して、商品に対する潜在的な購買層を推定し、推定した購買層に商品の広告を配信することが行われている。

【0003】

第一の消費者群の第一の行動に関する第一のデータと、第一消費者群及び第二消費者群の第二の行動に関する第二のデータと、に基づいて、第二の消費者群の第一の行動を推定する技術も既に知られている（特許文献1参照）。

【先行技術文献】

【特許文献】

【0004】

【文献】特開2020-161038号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

ところで、分析のために収集される人の行動に関するデータには、望まない情報の偏りが含まれ得る。例えば、分析に際して収集されるデータは、標本としての限られた人の集合の行動に関するデータである。

【0006】

限られた人の集合の行動は、母集団の行動に対して偏りを有する可能性がある。また、行動の観測環境を含む行動環境に起因する情報の偏りも存在し得る。従って、標本に基づいた分析により、人の行動に関する特徴を高精度に把握するためには、収集データに含まれる情報の偏りを補正できることが望ましい。

【0007】

そこで、本開示の一側面によれば、データに含まれる人の行動に関する情報の偏りを補正するのに適した情報処理技術を提供できることが望ましい。

【課題を解決するための手段】

【0008】

本開示の一側面によれば、情報処理システムが提供される。情報処理システムは、第一の取得部と、第二の取得部と、補正部とを備える。

第一の取得部は、第一の標本としての複数の第一の対象者に関する第一のデータセット

10

20

30

40

50

を取得する。第二の取得部は、第二の標本としての複数の第二の対象者に関する第二のデータセットを取得する。補正部は、第二のデータセットに基づき、第一のデータセットを補正する。

【0009】

第一のデータセットは、複数の第一の対象者のそれぞれに関して、対応する第一の対象者の行動を説明する行動データを、第一の行動データとして備える。第二のデータセットは、複数の第二の対象者のそれぞれに関して、対応する第二の対象者の行動を説明する行動データを、第二の行動データとして備える。

【0010】

第一の行動データは、複数の第一の行動に関して、第一の行動毎に、対応する第一の対象者の、対応する行動を数値により説明する。第二の行動データは、複数の第二の行動に関して、第二の行動毎に、対応する第二の対象者の、対応する行動を数値により説明する。複数の第二の行動には、複数の第一の行動が含まれる。

10

【0011】

補正部は、第二のデータセットに基づき、第一のデータセットにおける複数の第一の対象者の偏りに起因する数値のバイアスである第一のバイアス、及び、第一のデータセットにおける複数の第一の対象者のそれぞれの行動環境に起因する数値のバイアスである第二のバイアスを抑制する方向に、第一のデータセットにおける複数の第一の対象者に関する第一の行動毎の数値を補正することにより、第一のデータセットを補正する。

【0012】

人の行動に関するデータセットには、複数の要因によって情報の偏りが含まれ得る。例えば、人の行動は、人の属性の他、人が置かれている環境に依存すると考えられる。

20

【0013】

従って、上記第一のバイアス及び第二のバイアスを抑制する方向に、第一のデータセットを補正することによれば、第一の標本に関する第一のデータセットが有する複数のバイアスを適切に抑制することができる。従って、本開示の一側面によれば、複数の要因によって収集データに含まれる人の行動に関する情報の偏りを適切に補正することができる。

【0014】

本開示の一側面によれば、第一の標本は、特定グループに属する人に限定された人の集合であり得る。第二の標本は、特定グループに属さない人の集合又は特定グループに属する人に限定されない人の集合であり得る。第一の行動データと第二の行動データとの間では、行動の説明のために共通する変数である共通変数が用いられ得る。

30

【0015】

補正部は、共通変数で条件付けられる人が特定グループに所属する確率に対応する傾向スコアに基づき、第一の標本が特定グループに属する人の集合に限られていることに起因する第一のバイアスを抑制する方向に、第一のデータセットにおける複数の第一の対象者に関する第一の行動毎の数値を補正し得る。

【0016】

本開示の一側面によれば、第一の行動データ及び第二の行動データは、行動の主体である対象者の属性を、少なくとも一つの変数で表す属性情報を有し得る。第一の行動データと第二の行動データとの間では、属性の説明のために共通する変数である共通変数が用いられ得る。

40

【0017】

補正部は、この属性に関する共通変数を用いて、共通変数で条件付けられる人が特定グループに所属する確率に対応する傾向スコアに基づき、第一の標本が特定グループに属する人の集合に限られていることに起因する第一のバイアスを抑制する方向に、第一のデータセットにおける複数の第一の対象者に関する第一の行動毎の数値を補正し得る。

【0018】

本開示の一側面によれば、複数の第一の行動及び複数の第二の行動のそれぞれは、購買行動であり得る。第一の標本は、購買に関する特定のリワードプログラムに参加する消費

50

者、又は、特定の電子決済サービスを利用する消費者の集合であり得る。第一の行動データは、対応する第一の対象者からの識別情報の提供を通じて観測された対応する第一の対象者の購買行動を説明するデータであり得る。

【 0 0 1 9 】

行動に関するデータの収集に際しては、特定のリワードプログラム又は電子決済サービスを利用するケースが考えられるが、このケースでは、収集されるデータに消費者集合が限定されていること、及び、観測方式が限定されていることに伴うバイアスが含まれ得る。本開示の一側面によれば、このようなバイアスを含む収集データを適切に補正することが可能である。

【 0 0 2 0 】

本開示の一側面によれば、複数の第一の行動及び複数の第二の行動のそれぞれは、購買対象及び購買店舗の少なくとも一方によって区別される購買行動であり得る。本開示の一側面によれば、複数の第一の行動及び複数の第二の行動のそれぞれは、情報端末を通じた行動であり得る。

【 0 0 2 1 】

本開示の一側面によれば、補正部は、複数の地域に関して、地域毎に、複数の第一の行動のそれぞれの実行機会の大小を判別し、複数の第一の対象者のそれぞれの活動地域における実行機会の大小に起因する第二のバイアスを抑制する方向に、第一のデータセットにおける複数の第一の対象者に関する第一の行動毎の数値を補正し得る。

【 0 0 2 2 】

行動環境には地域による偏りがあり得る。行動機会の大小には、地域による偏りがあり得る。従って、実行機会の大小に起因するバイアスを抑制する方向に、行動に関するデータセットを補正することによれば、データセットに含まれる人の行動に関する情報の偏りを適切に補正することができる。

【 0 0 2 3 】

本開示の一側面によれば、複数の第一の行動及び複数の第二の行動のそれぞれは、少なくとも購買店舗によって区別される購買行動であって、複数の店舗のうちの一つの店舗における購買行動であり得る。一つの店舗には、同系の店舗群が含まれ得る。

【 0 0 2 4 】

補正部は、同系の店舗群の地理的分布に基づき、地理的分布に起因する第二のバイアスを抑制する方向に、第一のデータセットにおける複数の第一の対象者に関する第一の行動毎の数値を補正し得る。

【 0 0 2 5 】

同系の店舗群の地理的分布は、対応する店舗における購買機会の大小に影響を与え得る。従って、地理的分布に起因するバイアスを抑制するように、第一のデータセットを補正することは有意義である。

【 0 0 2 6 】

本開示の一側面によれば、補正部は、複数の店舗のそれぞれで利用される決済手段の特徴に基づき、複数の店舗間における決済手段の特徴の差異に起因する第二のバイアスを抑制する方向に、第一のデータセットにおける複数の第一の対象者に関する第一の行動毎の数値を補正し得る。

【 0 0 2 7 】

決済手段に応じて、購買行動の観測に偏りが生じ得る。従って、決済手段の特徴に基づき、第一のデータセットを補正することは有意義である。

【 0 0 2 8 】

本開示の一側面によれば、複数の第二の行動には、複数の第一の行動とは異なる複数の第三の行動が含まれ得る。第二の行動データは、第一の行動データにおいて説明されない複数の第三の行動に関して、第三の行動毎に、対応する第二の対象者の、対応する行動を数値により説明するデータであり得る。

【 0 0 2 9 】

10

20

30

40

50

この場合、補正部は、第二のデータセットに基づき、複数の第一の行動に関する数値から複数の第三の行動に関する数値の推定値を算出するための推定モデルを構築し得る。補正部は、複数の第一の対象者のそれぞれに関して、推定モデルに基づいて算出した、対応する第一の対象者の複数の第三の行動に関する推定値を、第一の行動データに追加することにより、第一のデータセットに欠損する複数の第三の行動に関する情報を補完するように、第一のデータセットを補正し得る。

【0030】

本開示の一側面によれば、推定モデルに基づく推定値の算出が、第一のバイアス及び第二のバイアスを抑制する方向に補正された、第一の行動データにおける複数の第一の行動に関する数値に基づいて行われ得る。補正された数値に基づく推定値の算出によれば、信頼性の高い第三の行動に関する情報を、第一のデータセットに追加することができ、第一のデータセットの情報価値を高めることができる。

10

【0031】

本開示の一側面によれば、上述した情報処理システムが有する機能の少なくとも一部を、コンピュータに実現させるためのコンピュータプログラムが提供され得る。本開示の一側面によれば、上述した情報処理システムにおける第一の取得部と、第二の取得部と、補正部としての機能を、コンピュータに実現させるためのコンピュータプログラムが提供され得る。

【0032】

本開示の一側面によれば、情報処理方法が提供されてもよい。情報処理方法は、コンピュータにより実行され得る。本開示の一側面によれば、情報処理方法は、上述した情報処理システムが実行する処理の少なくとも一部を実行する方法であり得る。

20

【0033】

本開示の一側面によれば、情報処理方法は、第一の標本としての複数の第一の対象者に関する第一のデータセットを取得することと、第二の標本としての複数の第二の対象者に関する第二のデータセットを取得することと、第二のデータセットに基づき、第一のデータセットを補正することとを含み得る。

【0034】

第一のデータセットは、複数の第一の対象者のそれぞれに関して、対応する第一の対象者の行動を説明する行動データを、第一の行動データとして備える。第二のデータセットは、複数の第二の対象者のそれぞれに関して、対応する第二の対象者の行動を説明する行動データを、第二の行動データとして備える。

30

【0035】

第一の行動データは、複数の第一の行動に関して、第一の行動毎に、対応する第一の対象者の、対応する行動を数値により説明する。第二の行動データは、複数の第二の行動に関して、第二の行動毎に、対応する第二の対象者の、対応する行動を数値により説明する。複数の第二の行動には、複数の第一の行動が含まれる。

【0036】

補正することは、第二のデータセットに基づき、第一のデータセットにおける複数の第一の対象者の偏りに起因する数値のバイアスである第一のバイアス、及び、第一のデータセットにおける複数の第一の対象者のそれぞれの行動環境に起因する数値のバイアスである第二のバイアスを抑制する方向に、第一のデータセットにおける複数の第一の対象者に関する第一の行動毎の数値を補正することにより、第一のデータセットを補正することを含み得る。この情報処理方法によれば、上述の情報処理システムと同様の効果を奏し得る。

40

【図面の簡単な説明】

【0037】

【図1】情報処理システムにおいて実現される機能を説明するブロック図である。

【図2】情報処理システムの電氣的構成を説明するブロック図である。

【図3】第一のデータセットの構成を説明する図である。

【図4】第二のデータセットの構成を説明する図である。

50

【図５】補完された第一のデータセットの構成を説明する図である。

【図６】情報処理システムのプロセッサが実行する補正補完処理を表すフローチャート（その１）である。

【図７】情報処理システムのプロセッサが実行する補正補完処理を表すフローチャート（その２）である。

【発明を実施するための形態】

【００３８】

以下に本開示の例示的实施形態を、図面を参照しながら説明する。

〔第一実施形態〕

図１に示す本実施形態の情報処理システム１は、第一グループの人の行動を説明する第一のデータセットＤＳ１を、第二グループの人の行動を説明する第二のデータセットＤＳ２を用いて補正するように構成される。情報処理システム１は更に、補正された第一のデータセットＤＳ１に欠損データを補完し、補完済データセットＤＣを出力するように構成される。

10

【００３９】

図２に示すように、情報処理システム１は、プロセッサ１１と、メモリ１２と、ストレージ１３と、ユーザインタフェース１５と、入出力インタフェース１７と、通信デバイス１９とを備える。

【００４０】

プロセッサ１１は、上記補正及び補完のために、ストレージ１３が記憶するコンピュータプログラムに従う処理を実行するように構成される。メモリ１２は、ＲＡＭを含み、プロセッサ１１による処理実行時に、作業領域として使用される。メモリ１２は、ストレージ１３から読み出されたコンピュータプログラム及びデータを一時記憶する。

20

【００４１】

ストレージ１３は、コンピュータプログラム、及び、上記処理に使用されるデータを記憶するように構成される。ストレージ１３の例には、ハードディスクドライブ及びソリッドステートドライブが含まれる。

【００４２】

ユーザインタフェース１５は、ユーザに向けて各種情報を表示するように構成される。ユーザインタフェース１５は更に、ユーザからの操作を受け付けて、対応する操作信号をプロセッサ１１に入力するように構成される。

30

【００４３】

ユーザインタフェース１５は、ユーザに向けて各種情報を表示するために、例えば液晶ディスプレイを備えることができる。ユーザインタフェース１５は、ユーザからの操作を受け付けるために、例えばキーボード及びポインティングデバイスを備えることができる。

【００４４】

入出力インタフェース１７は、外部装置を接続可能に構成される。入出力インタフェース１７は、外部記憶装置が接続された場合、プロセッサ１１からの指令に基づき、外部記憶装置からデータを読み出し可能であり、更には、外部記憶装置へデータを書き込み可能であるように構成される。入出力インタフェース１７は、例えばユニバーサルシリアルバス（ＵＳＢ）インタフェースであり得る。

40

【００４５】

通信デバイス１９は、通信ネットワークとして、ローカルエリアネットワーク及び／又は広域ネットワークに接続され、プロセッサ１１により制御されて、通信ネットワーク内の通信機器と通信するように構成される。

【００４６】

情報処理システム１のプロセッサ１１は、ストレージ１３が記憶するコンピュータプログラムに従う処理を実行することにより、データ取得部１１１、バイアス補正部１１３、モデル構築部１１５、欠損補完部１１７、及び出力部１１９として機能する。

【００４７】

50

データ取得部 111 は、第一のデータセット D S 1 として、上記補正及び補完対象のデータセットを取得する。データ取得部 111 は更に、第二のデータセット D S 2 として、補正及び補完に際して参照する参照用のデータセットを取得する。

【0048】

第一のデータセット D S 1 及び第二のデータセット D S 2 は、入出力インタフェース 17 を通じて外部記憶装置から取得され得る。あるいは、第一のデータセット D S 1 及び第二のデータセット D S 2 は、通信デバイス 19 を通じて外部の通信機器から取得され得る。

【0049】

本実施形態によれば、第一のデータセット D S 1 は、第一グループの消費者の購買行動を説明するデータセットである。第二のデータセット D S 2 は、第二グループの消費者の購買行動を説明するデータセットである。

10

【0050】

第一グループは、特定の消費者グループに対応する。具体的には、第一グループは、購買行動に応じたリワードが消費者に提供されるリワードプログラムに参加する消費者のグループである。

【0051】

リワードプログラムは、例えば、購買行動に応じたリワードとして、有価ポイントを消費者に提供するリワードプログラムである。消費者は、購買時に、リワードプログラムの会員 ID を、物理カード（所謂ポイントカード）又は情報端末の画面を通じて提示することにより、購買に応じた有価ポイントの提供を受けることができる。

20

【0052】

第二グループは、購買行動に関するパネル調査に参加する消費者のグループである。第二グループは、第一グループとは異なる消費者のグループであるが、第一グループと重複する消費者を部分的に含み得る。

【0053】

図 3 に例示される第一のデータセット D S 1 は、消費者毎のレコードを備える。第一のデータセット D S 1 における各レコードは、対応する消費者の属性データと、対応する消費者の第一の購買データと、を備える。

【0054】

以下では、第一のデータセット D S 1 が有する各レコードに対応する消費者のことを第一の消費者とも言う。第一の消費者の一群は、リワードプログラムに参加する消費者の一群である。

30

【0055】

属性データは、対応する第一の消費者の属性を説明する。属性は、デモグラフィック属性及びサイコグラフィック属性の少なくとも一方を含む。属性データは、例えば、対応する第一の消費者のデモグラフィック属性として、性別、年齢及び居住地域を説明する情報を有することができる。

【0056】

第一の購買データは、対応する第一の消費者の購買行動を説明する。第一の購買データは、特定の店舗グループに属する複数の店舗 H 1 , H 2 , ... 及び複数の商品 G 1 , G 2 , ... に関して、店舗及び商品の組合せ毎の特徴量 A を記述する。この特徴量 A は、購買対象としての商品、及び購買店舗の組み合わせによって区別される購買行動の特徴量 A である。例えば、第一の購買データは、店舗及び商品の組合せ毎に、対応する第一の消費者が、対応する店舗で、対応する商品を購入したか否かを表す特徴量 A を記述する。特徴量 A は、例えば数値 0 , 1 の二値で表される。

40

【0057】

特定の店舗グループは、リワードプログラムの加盟店舗の一群である。図 3 に示される店舗 H 1 , H 2 , ... は、リワードプログラムの加盟店舗 H 1 , H 2 , ... に対応する。すなわち、第一の購買データは、加盟店舗 H 1 , H 2 , ... の一群に関して、加盟店舗及び商品の組合せ毎の特徴量 A を記述する。

50

【 0 0 5 8 】

リワードプログラムに参加する第一の消費者は、加盟店舗において、会員IDを提示した上で、商品を購入することにより、その購買に応じたリワードの提供を受けることができる。加盟店舗及び非加盟店舗は、その店舗での購買によりリワードの提供を受けることができるか否かに応じて区別される店舗である。

【 0 0 5 9 】

第一のデータセットDS1は、消費者の購買行動のうち、消費者の識別情報である上記の会員IDの提示を伴って観測された消費者の購買行動のみを説明する。第一のデータセットDS1は、例えばリワードプログラムの運営会社から取得され得る。

【 0 0 6 0 】

図4に例示されるように、第二のデータセットDS2もまた、消費者毎のレコードを備える。第二のデータセットDS2における各レコードは、対応する消費者の属性データと、対応する消費者の第二の購買データと、を備える。

【 0 0 6 1 】

以下では、第二のデータセットDS2が有する各レコードに対応する消費者のことを第二の消費者とも言う。第二の消費者の一群は、購買行動に関するパネル調査に参加する消費者の一群である。

【 0 0 6 2 】

第二の消費者の一群は、第一の消費者の一群に属する消費者を含むかもしれないが、第一の消費者の一群に属さない消費者を含む、第一の消費者の一群とは異なる消費者の一群である。第二の消費者の一群は、第一の消費者を全く含まない消費者の一群であってもよい。

【 0 0 6 3 】

第二のデータセットDS2の属性データは、第一のデータセットDS1と同様に、対応する第二の消費者の属性を説明する。属性データは、例えば、対応する第二の消費者のデモグラフィック属性として、性別、年齢及び居住地域を説明する情報を有する。

【 0 0 6 4 】

図3及び図4において示される変数Xは、消費者属性を示し、第一のデータセットDS1及び第二のデータセットDS2において共通して用いられる。すなわち、第二のデータセットDS2は、第一のデータセットDS1と共通する変数Xを少なくとも部分的に用いて、第二の消費者のそれぞれの属性を説明する。共通する変数Xは、消費者の性別、年齢及び居住地域であり得る。以下において、変数Xが示す消費者属性のことを、消費者属性Xとも表現する。

【 0 0 6 5 】

第二の購買データは、対応する第二の消費者の購買行動を説明するデータである。第二の購買データは、上記特定の店舗グループに限定されない複数の店舗H1, H2, ..., L1, L2, ...及び複数の商品G1, G2, ...に関して、店舗及び商品の組合せ毎の購買行動に関する特徴量Aを記述する。

【 0 0 6 6 】

特定店舗グループに限定されない複数の店舗H1, H2, ..., L1, L2, ...は、加盟店舗H1, H2, ...及び非加盟店舗L1, L2, ...を含む複数の店舗H1, H2, ..., L1, L2, ...である。すなわち、第二の購買データは、第一の購買データが特徴量Aを記述する店舗及び商品の組み合わせの集合を含む、より大きな店舗及び商品の組合せの集合に関して、店舗及び商品の組合せ毎の特徴量Aを記述する。

【 0 0 6 7 】

第二の購買データは、第一の購買データと同様に、店舗及び商品の組合せ毎に、対応する第二の消費者が、対応する店舗で、対応する商品を購入したか否かを表す特徴量Aを記述するように構成され得る。特徴量Aは、例えば数値0, 1の二値で表される。

【 0 0 6 8 】

バイアス補正部113は、データ取得部111が取得した第一のデータセットDS1及

10

20

30

40

50

び第二のデータセット D S 2 を補正するように構成される。第一及び第二のデータセット D S 1 , D S 2 において購買行動が説明される第一及び第二の消費者の一群は、母集団である商品を購入する可能性のある市場全体の消費者の集合から抽出された第一及び第二の標本である。

【 0 0 6 9 】

第一のデータセット D S 1 では、購買行動が説明される第一の消費者の集合が、母集団のうち、リワードプログラムの参加者集合に限定されている。第二のデータセット D S 2 では、購買行動が説明される第二の消費者の集合が、母集団のうち、パネル調査の参加者集合に限定されている。

【 0 0 7 0 】

このように、第一のデータセット D S 1 及び第二のデータセット D S 2 は、母集団から抽出された異なる標本の購買行動を説明する。このため、第一のデータセット D S 1 と第二のデータセット D S 2 との間で単純に購買行動の特徴量 A を比較することは、適切ではない。

【 0 0 7 1 】

すなわち、第一のデータセット D S 1 及び第二のデータセット D S 2 は、第一の消費者集合と第二の消費者集合との間の偏りに応じた特徴量 A のバイアスを有する。以下、消費者集合の偏りに応じた特徴量 A のバイアスのことを第一のバイアスという。

【 0 0 7 2 】

更に言えば、本実施形態では、一つの特徴量 A が、同じ系列の複数の系列店舗における購買行動をまとめて表す。コンビニエンスストアやドラッグストア等に関しては、同じ系列（チェーン）に属する複数の系列店舗が存在する。一つの特徴量 A は、同じ系列に属する複数の系列店舗を一つの店舗とみなして、対応する店舗における消費者の購買行動を表す。

【 0 0 7 3 】

この場合、一つの特徴量 A に対応する一つの店舗には、対応する複数の系列店舗の地理的分布が存在することになり、この地理的分布は、地域によって異なる。換言すれば、各消費者の各店舗を利用した商品の購買行動に関する環境は異なる。従って、第一のデータセット D S 1 及び第二のデータセット D S 2 は、各消費者の上記環境に応じた特徴量 A の第二のバイアスを有する。

【 0 0 7 4 】

例えば、第一の地域には、第一系列のドラッグストアが 5 割、第二系列のドラッグストアが 3 割、第三系列のドラッグストアが 2 割存在するのにも関わらず、第二の地域には、第一系列のドラッグストアが 2 割、第二系列のドラッグストアが 3 割、第三系列のドラッグストアが 5 割存在することということが考えられる。この場合、各系列のドラッグストアでの購買環境及び購買機会は、第一の地域と第二の地域との間で異なる。

【 0 0 7 5 】

このことから、第一のデータセット D S 1 及び第二のデータセット D S 2 において、各レコードに記述される、対応する消費者の店舗及び商品の組合せ毎の特徴量 $A(i, j)$ は、真値 $q(i, j)$ に、第一のバイアス $p(i | x)$ と、第二のバイアス $p(i, j | z)$ と、が付加された観測値 $p(i, j) = q(i, j) \cdot p(i | x) \cdot p(i, j | z)$ に対応する特徴量 $A(i, j) = p(i, j)$ であると言える。

【 0 0 7 6 】

i は、消費者のインデックスであり、 j は、店舗及び商品の組合せのインデックスである。 $p(i | x)$ は、インデックス i に対応する消費者の購買行動の観測値 $p(i, j)$ に含まれる第一のバイアスを表し、値 $p(i, j | z)$ は、インデックス i に対応する消費者のインデックス j に対応する店舗及び商品の組合せの購買行動の観測値 $p(i, j)$ に含まれる第二のバイアスを表す。

【 0 0 7 7 】

特徴量 $A(i, j)$ は、インデックス i に対応する消費者が、インデックス j に対応す

10

20

30

40

50

る店舗で、インデックス j に対応する商品を購入する購買行動に関する特徴量 A 、例えば、対応する商品を購入したか否かを表す特徴量 A を表す。同様に、真値 $q(i, j)$ 及び観測値 $p(i, j)$ は、それぞれ順に、インデックス i に対応する消費者が、インデックス j に対応する店舗で、インデックス j に対応する商品を購入する購買行動に関する特徴量 A の真値及び観測値を表す。

【0078】

バイアス補正部 113 は、第一のデータセット $DS1$ 及び第二のデータセット $DS2$ に基づいて、第一のデータセット $DS1$ 及び第二のデータセット $DS2$ の各特徴量 $A(i, j)$ に含まれる第一のバイアス $p(i|x)$ 及び第二のバイアス $p(i, j|z)$ を推定し、特徴量 $A(i, j)$ を、第一のバイアス $p(i|x)$ 及び第二のバイアス $p(i, j|z)$ の逆数で、値 $C(i, j) = A(i, j) / (p(i|x) \cdot p(i, j|z))$ に補正する。これにより、バイアス補正部 113 は、特徴量 $A(i, j) = p(i, j)$ を真値 $q(i, j)$ に対応する値 $C(i, j)$ に補正する。

10

【0079】

モデル構築部 115 は、第二のデータセット $DS2$ が説明する加盟店舗 $H1, H2, \dots$ 及び非加盟店舗 $L1, L2, \dots$ を含む複数の店舗 $H1, H2, \dots, L1, L2, \dots$ における第二の消費者の購買行動に関する特徴量 $A(i = i2, j)$ を補正した補正後の特徴量 $C(i = i2, j)$ に基づいて、非加盟店舗における第一の消費者の購買行動に関する特徴量 $C(i = i1, j = j2)$ を、第一の消費者の加盟店舗における特徴量 $C(i = i1, j = j1)$ から推定するための推定モデルを構築する。

20

【0080】

ここでインデックス $i1$ は、第一の消費者のインデックスであり、インデックス $i2$ は、第二の消費者のインデックスである。インデックス $j1$ は、加盟店舗と商品との組合せインデックスであり、インデックス $j2$ は、非加盟店舗と商品との組合せのインデックスである。

【0081】

欠損補完部 117 は、第一の消費者毎に、対応する第一の消費者の第一のデータセット $DS1$ から特定される第一の消費者の加盟店舗における購買行動の特徴量 $C(i1, j1)$ を、推定モデルに入力して、対応する第一の消費者の非加盟店舗における購買行動の特徴量 $C(i1, j2)$ を推定する。

30

【0082】

欠損補完部 117 は、補正後の第一のデータセット $DS1$ に含まれる各レコードに、欠損データとして、対応する第一の消費者の非加盟店舗における購買行動を説明する推定購買データを追加することにより、当該レコードに欠損する非加盟店舗の購買データを補完する。

【0083】

各レコードに付加される推定購買データは、非加盟店舗及び商品の組合せ毎に、対応する第一の消費者の、対応する非加盟店舗及び商品の組合せに関する購買行動の特徴量 $C(i1, j2)$ を記述する。

【0084】

図5は、補正後の第一のデータセット $DS1$ に非加盟店舗の推定購買データを補完して構成される補完済データセット DC を例示する。図5によれば、補完済データセット DC は、加盟店舗 $H1, H2, \dots$ に関する特徴量 $C(i1, j1)$ を説明する購買データだけでなく、非加盟店舗 $L1, L2, \dots$ に関する特徴量 $C(i1, j2)$ を説明する購買データを有する。

40

【0085】

出力部 119 は、欠損補完部 117 により生成された補完済データセット DC を出力する。例えば、出力部 119 は、ストレージ 13 に補完済データセット DC を記録する。あるいは、出力部 119 は、補完済データセット DC を、ユーザインタフェース 15 を通じて、情報処理システム 1 のユーザに表示する。

50

【0086】

このようにして本実施形態の情報処理システム1は、第一のデータセットDS1に対するバイアス補正及び欠損補完を実行し、補完済データセットDCを出力する。

【0087】

続いて、プロセッサ11が、ユーザからの指示に基づいて実行する補正補完処理の詳細を、図6及び図7を用いて説明する。プロセッサ11は、補正補完処理を実行することにより、上述したデータ取得部111、バイアス補正部113、モデル構築部115、欠損補完部117、及び出力部119として機能する。

【0088】

図6に示す補正補完処理を開始すると、プロセッサ11は、補正及び補完対象の第一のデータセットDS1を取得し(S110)、更には、参照用の第二のデータセットDS2を取得する(S120)。

10

【0089】

その後、プロセッサ11は、第一のデータセットDS1及び第二のデータセットDS2に基づいて、消費者属性Xで条件付けられる消費者がリワードプログラムの参加者である確率に対応する傾向スコアを算出することを通じて、第一のデータセットDS1及び第二のデータセットDS2の各特徴量A(i, j)に含まれる第一のバイアスp(i | x)を推定する(S130)。

【0090】

すなわち、プロセッサ11は、消費者属性 $X = X_i$ を有する消費者iに関して、 $p(i | x_i) = p(i | e_i(X_i))$ となる傾向スコア $e_i(X_i) = p(x_i = 1 | X_i)$ を算出する。ここで、 x_i は、上記xに対応し、消費者iが、リワードプログラムの参加者であるか否かを表す。具体的に、 $x_i = 1$ は、消費者iは、リワードプログラムの参加者であることを示し、 $x_i = 0$ は、消費者iは、リワードプログラムの参加者ではないことを示す。第一の消費者は、 $x_i = 1$ に対応し、第二の消費者は、基本的に $x_i = 0$ に対応する。

20

【0091】

$p(x_i = 1 | X_i)$ は、消費者属性 $X = X_i$ を有する消費者iが、リワードプログラム参加者($x_i = 1$)である確率に対応する。 $p(i |) = p(i | e_i(X_i))$ は、傾向スコア $e_i(X_i)$ の消費者iの特徴量A(i, j)に含まれる第一のバイアスp(i |)であり、傾向スコアが値であるときに購買が観測される確率に対応する。

30

【0092】

リワードプログラムの参加者が加盟店舗において購買行動を行った事実がある場合に、その購買行動が観測される確率がであるときには、第一のバイアスp(i |)は、 \cdot に対応する。 $= 1$ であると仮定すれば、第一のバイアスp(i |) = である。 $= 1$ は、消費者が忘れずに会員IDを提示する場合に対応する。

【0093】

続くS140において、プロセッサ11は、第一のデータセットDS1及び第二のデータセットDS2における各特徴量A(i, j)に含まれる第二のバイアスp(i, j | z)を推定する。ここで、第二のバイアスp(i, j | z)の推定手法の詳細を説明する。

【0094】

上述した通り、本実施形態における第一のデータセットDS1及び第二のデータセットDS2は、店舗と商品との組み合わせ毎に特徴量A(i, j)を記述するが、店舗は、系列で区別され、同系列の系列店舗群は、一つの店舗として取り扱われる。

40

【0095】

このため、消費者iのインデックスjに対応する店舗h[j]及び商品の組合せに関する購買行動の特徴量A(i, j)には、消費者iの居住地域r[i]における店舗h[j]の、対応する店舗カテゴリg[j]内でのシェア $z_{r[i]g[j]}$ に伴うバイアスp(i, j | $z_{r[i]g[j]}$)が発生していると言うことができる。

【0096】

ここでいう店舗カテゴリは、ドラッグストア及びコンビニエンスストアなどの業態に対

50

応する。店舗カテゴリが、ドラッグストアであるときには、居住地域におけるドラッグストアのうち、注目する系列のドラッグストアのシェアが、ここでいう店舗カテゴリ内のシェアに対応する。本実施形態では、以下、店舗カテゴリ内のシェアのことを、単にシェアという。

【0097】

例えば、消費者の居住地域におけるドラッグストアのシェアのうち、第一系列のドラッグストアのシェアが第二系列のドラッグストアのシェアよりも小さい場合には、第一系列のドラッグストアのシェアが第二系列のドラッグストアのシェアより大きい場合と比較して、消費者の第一系列のドラッグストアにおける購買行動の実行機会が小さくなり、第一のデータセットDS1において説明される第一系列のドラッグストアの購買行動の観測量は小さくなる。

10

【0098】

このため、プロセッサ11は、各特微量 $A(i, j)$ に関して、対応する消費者の居住地域 $r[i]$ における店舗 $h[j]$ のシェア $z_{r[i]g[j]}$ を算出し、そのシェア $z_{r[i]g[j]}$ に基づいて、第二のバイアス $p(i, j | z_{r[i]g[j]})$ を推定する。

【0099】

具体的には、シェア $z_{r[i]g[j]}$ を、第二のバイアス $p(i, j | z_{r[i]g[j]}) = z_{r[i]g[j]}$ であると推定する。シェア $z_{r[i]g[j]}$ は、地域 $r[i]$ における店舗 $h[j]$ を通じた消費者の購買行動の実行機会の大小を表すと言える。

【0100】

20

シェア $z_{r[i]g[j]}$ の算出のために、ストレージ13は、地域及び店舗カテゴリ毎に、対応する地域における、対応する店舗カテゴリの店舗数合計 M と、対応する店舗カテゴリにおける各系列の系列店舗数 N とを記述するテーブルを記憶しておくことができる。

【0101】

プロセッサ11は、店舗 $h[j]$ のシェア $z_{r[i]g[j]}$ を、居住地域 $r[i]$ に対応する地域における店舗 $h[j]$ に対応する店舗カテゴリ $g[j]$ の店舗 $h[j]$ に対応する系列店舗数 N と、店舗カテゴリ $g[j]$ の店舗数合計 M とを、用いて、式 $z_{r[i]g[j]} = N / M$ に従って算出することができる。

【0102】

プロセッサ11は、このようにして、S130, S140で第一のデータセットDS1及び第二のデータセットDS2における各特微量 $A(i, j)$ に含まれる第一のバイアス $p(i | x)$ 及び第二のバイアス $p(i, j | z)$ を推定後、S150の処理を実行する。

30

【0103】

S150において、プロセッサ11は、第一のデータセットDS1及び第二のデータセットDS2が説明する各特微量 $A(i, j)$ を、第一のバイアス $p(i | x)$ 及び第二のバイアス $p(i, j | z)$ を抑制する方向に補正する。

【0104】

すなわち、プロセッサ11は、各特微量 $A(i, j)$ に、対応する消費者 i の属性に起因した第一のバイアス $p(i | x)$ の逆数を作用させると共に、対応する消費者 i の居住地域 $r[i]$ における店舗 $h[j]$ のシェアに起因した第二のバイアス $p(i, j | z)$ の逆数を作用させることにより、特微量 $A(i, j)$ に対応する補正值 $C(i, j) = A(i, j) / (p(i | x) \cdot p(i, j | z))$ を算出する。居住地域は、対応する消費者の活動地域に対応する。

40

【0105】

続くS160において、プロセッサ11は、第二のデータセットDS2の各レコードが記述する特微量 $A(i_2, j)$ のそれぞれを値 $C(i_2, j)$ に補正した補正後の第二のデータセットDS2に基づいて、非加盟店舗における第一の消費者の購買行動に関する特微量 $C(i_1, j_2)$ を、第一の消費者の加盟店舗における特微量 $C(i_1, j_1)$ から推定するための推定モデルを構築する。

【0106】

50

補正後の第二のデータセットDS2は、消費者i2毎に、消費者i2の消費者属性Xと、対応する消費者i2の各加盟店舗における各商品の購買行動に関する特徴量C(i2, j1)と、対応する消費者i2の各非加盟店舗における各商品の購買行動に関する特徴量C(i2, j2)とを、記述する。

【0107】

従って、補正後の第二のデータセットDS2を教師データとして用いれば、例えば機械学習により、消費者iの消費者属性 $X = X_i$ と、消費者iの加盟店舗の購買行動に関する特徴量C(i, j1)と、から、消費者iの非加盟店舗の購買行動に関する特徴量C(i, j2)を推定するための推定モデルを構築することができる。推定モデルは、例えば、回帰により構築されてもよいし、ニューラルネットワークを用いて構築されてもよいし、その他の機械学習モデルを用いて構築されてもよい。

10

【0108】

続くS170において、プロセッサ11は、補正後の第一のデータセットDS1と上記推定モデルとを用いて、第一の消費者i1のそれぞれの、非加盟店舗のそれぞれにおける店舗及び商品の組合せ毎の特徴量C(i1, j2)を算出する。

【0109】

具体的には、プロセッサ11は、補正後の第一のデータセットDS1のレコード毎に、対応するレコードが示す消費者の消費者属性 $X = X_i$ と、加盟店舗における店舗及び商品の組合せ毎の購買行動に関する特徴量C(i1, j1)と、を推定モデルに入力し、推定モデルから、出力として、対応する消費者の非加盟店舗における店舗及び商品の組合せ毎の購買行動に関する特徴量C(i1, j2)を取得する。

20

【0110】

その後、プロセッサ11は、補正後の第一のデータセットDS1の各レコードに、対応する消費者の非加盟店舗における店舗及び商品の組合せ毎の購買行動に関する特徴量C(i1, j2)を記述した推定購買データを追加することにより、第一のデータセットDS1に対する欠損補完を実現する(S180)。

【0111】

S180において、プロセッサ11は、補正後の第一のデータセットDS1の各レコードに対する推定購買データの追加により、第一のデータセットDS1に対応する補完済データセットDCを生成する。その後、プロセッサ11は、上記生成した補完済データセットDCを出力し(S190)、当該補正補完処理を終了する。

30

【0112】

以上に説明した本実施形態の情報処理システム1によれば、消費者の購買行動に関する第一のデータセットDS1を、消費者の購買行動を説明する特徴量Aに含まれる複数のバイアスを適切に取り除くように補正することができる。更には、バイアスを取り除いた特徴量Cに基づいて、第一のデータセットDS1に欠損するデータを、第二のデータセットDS2を用いて補完する。

【0113】

従って、本実施形態によれば、情報の偏りを含む購買行動に関するデータセットに対し、バイアス補正及び欠損補完を良好に行うことができ、データセットの情報確度を高めることができる。具体的には、リワードプログラムを利用して収集した購買行動に関するデータセットに含まれる、その収集形態に応じた情報の偏りを抑制することができ、更には、リワードプログラムを通じたデータ収集では得られない非加盟店舗の購買データを補完できることから、消費者の購買行動の分析に役立つデータセットを生成することができる。

40

【0114】

[第二実施形態]

続いて第二実施形態の情報処理システム1を説明する。但し、第二実施形態の情報処理システム1は、第一のデータセットDS1及び第二のデータセットDS2の内容が第一実施形態とは異なり、それにより、バイアス補正部113、モデル構築部115、及び欠損

50

補完部 1 1 7 により実現される機能の一部が第一実施形態と異なるだけである。従って、以下では、第二実施形態の情報処理システム 1 の構成として、第一実施形態とは異なる構成を選択的に説明する。

【 0 1 1 5 】

本実施形態のデータ取得部 1 1 1 (図 1 参照) が取得する第一のデータセット D S 1 は、第一グループの人の行動として、リワードプログラム参加者の購買行動ではなく、特定の電子決済サービスを利用する消費者の購買行動を説明するデータセットである。

【 0 1 1 6 】

具体的には、第一のデータセット D S 1 は、クレジットカード利用者に対応する消費者の、クレジットカード決済可能店舗における購買行動を説明するデータセットである。第一のデータセット D S 1 は、例えばクレジットカードの決済記録から作成される。クレジットカード利用者は、特定企業から発行されるクレジットカードの名義人に対応する消費者であり得る。

10

【 0 1 1 7 】

具体的に、第一のデータセット D S 1 (図 3 参照) は、クレジットカード利用者に対応する消費者毎に、対応する消費者のクレジットカード決済による購買行動を説明するレコードを有する。各レコードは、対応する消費者のクレジットカードを用いた購買行動に関する特徴量 A として、クレジットカード決済可能店舗に該当する複数の店舗 H 1 , H 2 , ... に関する店舗及び商品の組合せ毎の特徴量 A を、消費者属性 X と共に記述する。

【 0 1 1 8 】

20

一方、データ取得部 1 1 1 が取得する第二のデータセット D S 2 (図 4 参照) は、パネル調査の参加者毎に、対応する消費者の決済手段によらない購買行動を説明するレコードを有する。各レコードは、対応する消費者の購買行動に関する特徴量 A として、クレジットカード決済可能店舗及びクレジットカード決済不可能店舗を含む複数の店舗 H 1 , H 2 , ... , L 1 , L 2 , ... に関する店舗及び商品の組合せ毎の購買行動に関する特徴量 A を、消費者属性 X と共に記述する。

【 0 1 1 9 】

本実施形態によれば、第一実施形態と同様に、第一のデータセット D S 1 及び第二のデータセット D S 2 が説明する各特徴量 A (i , j) には、第一のデータセット D S 1 と第二のデータセット D S 2 との間の消費者集合の相違に応じた第一のバイアス p (i | x) が存在する。

30

【 0 1 2 0 】

この他、消費者による決済手段の選択、すなわちクレジットカード決済を選択するか否かは、店舗によって異なる。従って、特徴量 A (i , j) には、消費者及び店舗に応じた第二のバイアス p (i , j | z) が存在する。

【 0 1 2 1 】

このため、バイアス補正部 1 1 3 は、第一実施形態と同様に、傾向スコア を用いて、第一のバイアス p (i | x) を算出する。バイアス補正部 1 1 3 は、傾向スコア $e_i(X_i) = p(x_i = 1 | X_i)$ として、消費者属性 $X = X_i$ を有する消費者 i が、クレジットカード利用者 ($x_i = 1$) である確率 $p(x_i = 1 | X_i)$ を算出することができ、これにより、第一のバイアス $p(i |) =$ を推定することができる。

40

【 0 1 2 2 】

バイアス補正部 1 1 3 は更に、第二のバイアス p (i , j | z) として、対応する消費者 i の、インデックス j に対応する店舗におけるクレジットカード利用率 V (i , j) を算出する ($p(i , j | z) = V(i , j)$) 。

【 0 1 2 3 】

ある店舗におけるクレジットカード利用率は、当該店舗における決済手段を限定しない決済全体のうち、決済手段としてクレジットカードが利用された決済の割合を意味する。クレジットカード利用率は、対応する店舗における決済手段の利用に関する特徴を表す。すなわち、本実施形態における第二のバイアス p (i , j | z) は、各店舗における決定

50

手段の特徴の差異に起因するバイアスである。

【0124】

クレジットカード利用率の算出のために、各店舗におけるクレジットカードの決済データと、各店舗の販売記録であるPOS(Point Of Sales)データと、が活用され得る。クレジットカード利用率の算出のために、パネル調査参加者のID-POSデータが活用されてもよい。

【0125】

ID-POSデータに基づけば、消費者属性から、対応する消費者属性を有する消費者の各店舗におけるクレジットカード利用率 V を推定する推定モデルが構築され得る。構築された推定モデルに基づいて、第一のデータセットDS1に対応する各消費者の各店舗におけるクレジットカード利用率 $V(i, j)$ が推定され得る。

10

【0126】

バイアス補正部113は、上記の第一のバイアス $p(i|z)$ 及び第二のバイアス $p(i, j|z) = V(i, j)$ を用いて、第一のデータセットDS1における各特徴量 $A(i, j)$ を、値 $C(i, j) = A(i, j) / (p(i|z) * p(i, j|z))$ で補正する。第二のデータセットDS2における各特徴量 $A(i, j)$ を、値 $C(i, j) = A(i, j) / p(i|z)$ で補正する。

【0127】

モデル構築部115は、このように補正された第二のデータセットDS2に基づいて、消費者属性 X 及びクレジットカード利用可能店舗の特徴量 $C(i, j1)$ から、クレジットカード利用不可能店舗の特徴量 $C(i, j2)$ を推定する推定モデルを構築する。

20

【0128】

欠損補完部117は、この推定モデルを用いて、第一のデータセットDS1の各レコードに対して、対応する消費者 $i1$ のクレジットカード利用不可能店舗の特徴量 $C(i1, j2)$ を記述した推定購買データを、欠損データとして追加することにより、第一のデータセットDS1に対する欠損補完を行う。出力部119は、このようにして生成された補完済データセットDC(図5参照)を出力する。

【0129】

本実施形態によれば、限定された決済手段を用いた決済記録に基づいて用意される、情報の偏りを含む購買行動に関するデータセットに対して、バイアス補正及び欠損補完を良好に行うことができ、データセットの情報確度を高めることができる。具体的には、クレジットカード決済の記録を利用して収集した購買行動に関するデータセットに含まれる、その収集形態に応じた情報の偏りを抑制することができ、更には、クレジットカード決済の記録を用いるだけでは得られないクレジットカード決済不可能店舗の購買データを補完できることから、消費者の購買行動の分析に役立つデータセットを生成することができる。

30

【0130】

[第三実施形態]

続いて第三実施形態の情報処理システム1を説明する。但し、第三実施形態の情報処理システム1は、第一のデータセットDS1及び第二のデータセットDS2の内容が第一実施形態とは異なり、それにより、バイアス補正部113、モデル構築部115、及び欠損補完部117により実現される機能の一部が第一実施形態と異なるだけである。従って、以下では、第三実施形態の情報処理システム1の構成として、第一実施形態とは異なる構成を選択的に説明する。

40

【0131】

本実施形態においてデータ取得部111が取得する第一のデータセットDS1は、特定のオンラインサービスを利用する情報端末のユーザ毎のレコードを有する。各レコードは、対応するユーザの、情報端末の利用に関する種類毎の特徴量 A を記述する。

【0132】

特定のオンラインサービスを利用するユーザは、例えば、特定のアプリケーションソフトウェアを利用して特定のオンラインサービス(例えばオンラインショッピングモール)

50

を利用するユーザであり得る。

【0133】

情報端末は、例えばスマートフォンやタブレット端末である。種類毎の特徴量 A は、具体的には、情報端末にインストールされるアプリケーションソフトウェア毎の利用に関する特徴量である。

【0134】

すなわち、第一のデータセット D S 1 の各レコードは、対応するユーザの情報端末におけるアプリケーションソフトウェア毎の利用に関する特徴量 A を記述する。特徴量 A は、上記特定のオンラインサービスの利用に際して使用される上記特定のアプリケーションソフトウェアを通じて計測され得る。

10

【0135】

第二のデータセット D S 2 は、パネル調査の参加者に対応する情報端末のユーザ毎のレコードを有する。各レコードは、対応するユーザの、情報端末の利用に関する種類毎の特徴量 A を記述する。具体的には、各レコードは、対応するユーザの情報端末におけるアプリケーションソフトウェア毎の利用に関する特徴量 A を記述する。第二のデータセット D S 2 に関して、各ユーザのアプリケーションソフトウェア毎の利用に関する特徴量 A は、パネル調査の参加者に対して配布される計測用のアプリケーションソフトウェアを通じて計測され得る。

【0136】

第一のデータセット D S 1 及び第二のデータセット D S 2 において記述される特徴量 A は、対応するアプリケーションソフトウェアを通じて利用されるオンラインサービスの利用に関する特徴量 A であり得る。この場合、特徴量 A は、ユーザのオンライン行動に関する特徴量といえる。

20

【0137】

バイアス補正部 113 は、第一実施形態と同様に、傾向スコア $= e_i(X_i) = p(x_i = 1 | X_i)$ として、消費者属性 $X = X_i$ を有するユーザ i が、上記特定のオンラインサービスの利用者 ($x_i = 1$) である確率 $p(x_i = 1 | X_i)$ を算出し、これにより、第一のバイアス $p(i |) =$ を推定する。

【0138】

バイアス補正部 113 は更に、第二のバイアス $p(i, j | z)$ として、対応するユーザ i が使用する情報端末のオペレーティングシステム (OS) における、インデックス j に対応するアプリケーションソフトウェアのカテゴリ内シェア $W(i, j)$ を算出する ($p(i, j | z) = W(i, j)$)。

30

【0139】

ある OS において注目するアプリケーションソフトウェアのカテゴリ内シェアは、当該 OS 上で利用される当該注目するアプリケーションソフトウェアの、利用量合計 (例えばインストール数又は利用時間合計) を、同じ OS 上で利用される当該注目するアプリケーションソフトウェアと同じカテゴリの全アプリケーションソフトウェアの利用量合計で除算して算出される値である。ここでいう利用量合計は、同じ OS を使用する複数ユーザの利用量合計である。各 OS におけるアプリケーションソフトウェア毎のカテゴリ内シェアは、例えば、パネル調査データを通じて判別される。

40

【0140】

各アプリケーションソフトウェアのカテゴリ内シェアは OS によって異なる。このため、各ユーザのアプリケーションソフトウェア毎の利用に関する特徴量 A には、ユーザが使用する情報端末の OS に応じた第二のバイアス $p(i, j | z)$ が生じ得る。

【0141】

このため、バイアス補正部 113 は、ユーザ i 毎に、対応するユーザ i のアプリケーションソフトウェア j 毎の特徴量 $A(i, j)$ を、対応するユーザ i の情報端末の OS における対応するアプリケーションソフトウェア j のカテゴリ内シェアで除算することにより、特徴量 $A(i, j)$ におけるシェアに起因した第二のバイアス $p(i, j | z)$ を取り

50

除く。

【0142】

具体的には、バイアス補正部113は、第一のデータセットDS1及び第二のデータセットDS2における各ユーザのアプリケーションソフトウェア毎の特徴量 $A(i, j)$ を、値 $C(i, j) = A(i, j) / (p(i |) * p(i, j | z))$ で補正する。

【0143】

付言すると、本実施形態において取り扱われるアプリケーションソフトウェアには、第一のデータセットDS1が特徴量Aを記述する第一のアプリケーションソフトウェアの他、第一のデータセットDS1では特徴量Aが記述されない第二のアプリケーションソフトウェアが含まれる。

【0144】

すなわち、第一のデータセットDS1は、第一のアプリケーションソフトウェアに関して、ユーザ毎及びアプリケーションソフトウェア毎の特徴量 $A(i, j)$ を記述する。第二のデータセットDS2は、第一及び第二のアプリケーションソフトウェアに関して、ユーザ毎及びアプリケーションソフトウェア毎の特徴量 $A(i, j)$ を記述する。

【0145】

モデル構築部115は、バイアス補正部113で補正された第二のデータセットDS2に基づいて、第一のアプリケーションソフトウェアに関するアプリケーション毎の特徴量 $C(i, j)$ と、対応するユーザiの消費者属性 $X = X_i$ と、から対応するユーザiの第二のアプリケーションソフトウェアに関するアプリケーション毎の特徴量 $C(i, j)$ を推定する推定モデルを構築する。

【0146】

欠損補完部117は、補正後の第一のデータセットDS1に、推定モデルを用いて推定された第二のアプリケーションソフトウェアに関するユーザ毎及びアプリケーションソフトウェア毎の特徴量 $C(i, j)$ を追加することにより、第一のデータセットDS1に対する欠損補完を実現する。出力部119は、このようにして生成された補完済データセットDCを出力する。

【0147】

本実施形態によれば、限定されたオンラインサービスの利用ユーザの行動記録に基づいて用意される、情報の偏りを含むアプリケーションソフトウェアの利用に関する特徴量を説明するデータセットに対して、バイアス補正及び欠損補完を良好に行うことができ、データセットの情報確度を高めることができる。具体的には、限定したユーザの情報端末を通じた行動記録を利用して収集したアプリケーションソフトウェアの利用に関する特徴量を説明するデータセットに含まれる、その収集形態に応じた情報の偏りを抑制することができ、更には、限定したユーザの情報端末から得られない、当該情報端末を通じた行動記録を補完できることから、消費者の情報端末を通じた行動の分析に役立つデータセットを生成することができる。

【0148】

[その他の実施形態]

本開示は、上述した第一から第三実施形態に限定されるものではなく、種々の態様を採ることができる。情報処理システム1が取り扱う第一のデータセットDS1及び第二のデータセットDS2の構成は、第一から第三実施形態に限定されない。情報処理システム1は、欠損補完を行わないように構成されてもよく、出力部119は、バイアス補正部113による補正後の第一のデータセットDS1を出力するように構成されてもよい。

【0149】

上記実施形態における1つの構成要素が有する機能は、複数の構成要素に分散して設けられてもよい。複数の構成要素が有する機能は、1つの構成要素に統合されてもよい。上記実施形態の構成の一部は、省略されてもよい。上記実施形態の構成の少なくとも一部は、他の上記実施形態の構成に対して付加又は置換されてもよい。特許請求の範囲に記載の文言から特定される技術思想に含まれるあらゆる態様が本開示の実施形態である。

10

20

30

40

50

【 0 1 5 0 】

[本明細書が開示する技術思想]

本明細書には、次の技術思想が開示されていると理解することができる。

[項目 1]

第一の標本としての複数の第一の対象者に関する第一のデータセットを取得する第一の取得部と、

第二の標本としての複数の第二の対象者に関する第二のデータセットを取得する第二の取得部と、

前記第二のデータセットに基づき、前記第一のデータセットを補正する補正部と、
を備え、

前記第一のデータセットは、前記複数の第一の対象者のそれぞれに関して、対応する第一の対象者の行動を説明する行動データを、第一の行動データとして備え、

前記第二のデータセットは、前記複数の第二の対象者のそれぞれに関して、対応する第二の対象者の行動を説明する行動データを、第二の行動データとして備え、

前記第一の行動データは、複数の第一の行動に関して、第一の行動毎に、前記対応する第一の対象者の、対応する行動を数値により説明し、

前記第二の行動データは、複数の第二の行動に関して、第二の行動毎に、前記対応する第二の対象者の、対応する行動を数値により説明し、

前記複数の第二の行動には、前記複数の第一の行動が含まれ、

前記補正部は、前記第一のデータセットにおける前記複数の第一の対象者の偏りに起因する前記数値のバイアスである第一のバイアス、及び、前記第一のデータセットにおける前記複数の第一の対象者のそれぞれの行動環境に起因する前記数値のバイアスである第二のバイアスを抑制する方向に、前記第一のデータセットにおける前記複数の第一の対象者に関する前記第一の行動毎の前記数値を補正することにより、前記第一のデータセットを補正する情報処理システム。

[項目 2]

前記第一の標本は、特定グループに属する人に限定された人の集合であり、

前記第二の標本は、前記特定グループに属さない人の集合又は前記特定グループに属する人に限定されない人の集合であり、

前記第一の行動データと前記第二の行動データとの間では、前記行動の説明のために共通する変数である共通変数が用いられ、

前記補正部は、前記第一のデータセット及び前記第二のデータセットに基づく前記共通変数で条件付けられる人が前記特定グループに所属する確率に対応する傾向スコアに基づき、前記第一の標本が前記特定グループに属する人の集合に限られていることに起因する前記第一のバイアスを抑制する方向に、前記第一のデータセットにおける前記複数の第一の対象者に関する前記第一の行動毎の前記数値を補正する項目 1 記載の情報処理システム。

[項目 3]

前記第一の標本は、特定グループに属する人に限定された人の集合であり、

前記第二の標本は、前記特定グループに属さない人の集合又は前記特定グループに属する人に限定されない人の集合であり、

前記第一の行動データ及び前記第二の行動データは、前記行動の主体である対象者の属性を、少なくとも一つの変数で表す属性情報を有し、

前記第一の行動データと前記第二の行動データとの間では、前記属性の説明のために共通する変数である共通変数が用いられ、

前記補正部は、前記第一のデータセット及び前記第二のデータセットに基づく前記共通変数で条件付けられる人が前記特定グループに所属する確率に対応する傾向スコアに基づき、前記第一の標本が前記特定グループに属する人の集合に限られていることに起因する前記第一のバイアスを抑制する方向に、前記第一のデータセットにおける前記複数の第一の対象者に関する前記第一の行動毎の前記数値を補正する項目 1 記載の情報処理システム。

[項目 4]

前記複数の第一の行動及び前記複数の第二の行動のそれぞれは、購買行動であり、

前記第一の標本は、購買に関する特定のリワードプログラムに参加する消費者、又は、特定の電子決済サービスを利用する消費者の集合であり、

前記第一の行動データは、前記対応する第一の対象者からの識別情報の提供を通じて観測された前記対応する第一の対象者の購買行動を説明するデータである項目1～項目3のいずれか一項記載の情報処理システム。

[項目 5]

前記複数の第一の行動及び前記複数の第二の行動のそれぞれは、購買対象及び購買店舗の少なくとも一方によって区別される購買行動である項目1～項目4のいずれか一項記載の情報処理システム。

10

[項目 6]

前記複数の第一の行動及び前記複数の第二の行動のそれぞれは、情報端末を通じた行動である項目1～項目3のいずれか一項記載の情報処理システム。

[項目 7]

前記補正部は、複数の地域に関して、地域毎に、前記複数の第一の行動のそれぞれの実行機会の大小を判別し、前記複数の第一の対象者のそれぞれの活動地域における前記実行機会の大小に起因する前記第二のバイアスを抑制する方向に、前記第一のデータセットにおける前記複数の第一の対象者に関する前記第一の行動毎の前記数値を補正する項目1～項目5のいずれか一項記載の情報処理システム。

[項目 8]

20

前記複数の第一の行動及び前記複数の第二の行動のそれぞれは、少なくとも購買店舗によって区別される購買行動であって、複数の店舗のうちの一つの店舗における購買行動であり、

前記一つの店舗には、同系の店舗群が含まれ、

前記補正部は、前記同系の店舗群の地理的分布に基づき、前記地理的分布に起因する前記第二のバイアスを抑制する方向に、前記第一のデータセットにおける前記複数の第一の対象者に関する前記第一の行動毎の前記数値を補正する項目1～項目5のいずれか一項記載の情報処理システム。

[項目 9]

前記複数の第一の行動及び前記複数の第二の行動のそれぞれは、少なくとも購買店舗によって区別される購買行動であって、複数の店舗のうちの一つの店舗における購買行動であり、

30

前記補正部は、前記複数の店舗のそれぞれで利用される決済手段の特徴に基づき、前記複数の店舗間における前記決済手段の特徴の差異に起因する前記第二のバイアスを抑制する方向に、前記第一のデータセットにおける前記複数の第一の対象者に関する前記第一の行動毎の前記数値を補正する項目1～項目5のいずれか一項記載の情報処理システム。

[項目 10]

前記複数の第二の行動には、前記複数の第一の行動とは異なる複数の第三の行動が含まれ、

前記第二の行動データは、前記第一の行動データにおいて説明されない前記複数の第三の行動に関して、前記第三の行動毎に、前記対応する第二の対象者の、対応する行動を数値により説明し、

40

前記補正部は、

前記第二のデータセットに基づき、前記複数の第一の行動に関する前記数値から前記複数の第三の行動に関する前記数値の推定値を算出するための推定モデルを構築し、

前記複数の第一の対象者のそれぞれに関して、前記推定モデルに基づいて算出した、前記対応する第一の対象者の前記複数の第三の行動に関する前記推定値を、前記第一の行動データに追加することにより、前記第一のデータセットに欠損する前記複数の第三の行動に関する情報を補完するように、前記第一のデータセットを補正し、

前記推定モデルに基づく前記推定値の算出は、前記第一のバイアス及び前記第二のバイ

50

アスを抑制する方向に補正された、前記第一の行動データにおける前記複数の第一の行動に関する前記数値に基づいて行われる項目 1 ～ 項目 9 のいずれか一項記載の情報処理システム。

[項目 1 1]

項目 1 ～ 項目 1 0 のいずれか一項記載の情報処理システムにおける前記第一の取得部と、前記第二の取得部と、前記補正部としての機能を、コンピュータに実現させるためのコンピュータプログラム。

[項目 1 2]

コンピュータにより実行される情報処理方法であって、

第一の標本としての複数の第一の対象者に関する第一のデータセットを取得することと、
第二の標本としての複数の第二の対象者に関する第二のデータセットを取得することと、
前記第二のデータセットに基づき、前記第一のデータセットを補正することと
を含み、

前記第一のデータセットは、前記複数の第一の対象者のそれぞれに関して、対応する第一の対象者の行動を説明する行動データを、第一の行動データとして備え、

前記第二のデータセットは、前記複数の第二の対象者のそれぞれに関して、対応する第二の対象者の行動を説明する行動データを、第二の行動データとして備え、

前記第一の行動データは、複数の第一の行動に関して、第一の行動毎に、前記対応する第一の対象者の、対応する行動を数値により説明し、

前記第二の行動データは、複数の第二の行動に関して、第二の行動毎に、前記対応する第二の対象者の、対応する行動を数値により説明し、

前記複数の第二の行動には、前記複数の第一の行動が含まれ、

前記補正することは、前記第一のデータセットにおける前記複数の第一の対象者の偏りに起因する前記数値のバイアスである第一のバイアス、及び、前記第一のデータセットにおける前記複数の第一の対象者のそれぞれの行動環境に起因する前記数値のバイアスである第二のバイアスを抑制する方向に、前記第一のデータセットにおける前記複数の第一の対象者に関する前記第一の行動毎の前記数値を補正することにより、前記第一のデータセットを補正することを含む情報処理方法。

【符号の説明】

【 0 1 5 1 】

1 ... 情報処理システム、 1 1 ... プロセッサ、 1 2 ... メモリ、 1 3 ... ストレージ、 1 5 ... ユーザインタフェース、 1 7 ... 入出力インタフェース、 1 9 ... 通信デバイス、 1 1 1 ... データ取得部、 1 1 3 ... バイアス補正部、 1 1 5 ... モデル構築部、 1 1 7 ... 欠損補完部、 1 1 9 ... 出力部。

10

20

30

40

50

【図面】

【図 1】

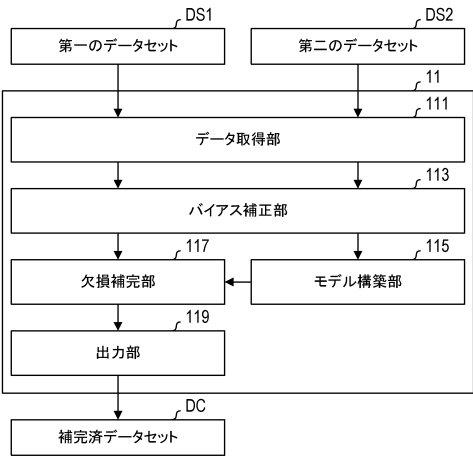


FIG. 1

【図 2】

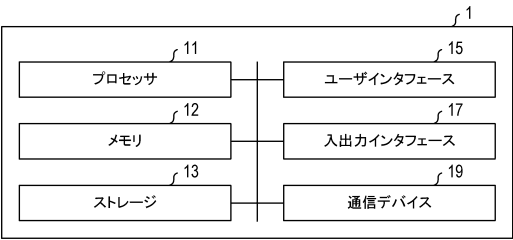


FIG. 2

10

【図 3】

消費者	属性	店舗H1		店舗H2	
		商品G1	商品G2	商品G1	商品G2
...
i1	X	A(1,1)	A(1,2)	A(1,1)	A(1,12)
...

FIG. 3

【図 4】

消費者	属性	店舗H1		店舗H2		店舗L1		店舗L2	
		商品G1	商品G2	商品G1	商品G2	商品G1	商品G2	商品G1	商品G2
...
i2	X	A(2,1)	A(2,2)	A(2,1)	A(2,12)	A(2,5)	A(2,52)
...

FIG. 4

20

30

40

50

【図 5】

消費者	属性	店舗+1		店舗+2		店舗+1		店舗+2		DC
		商品G1	商品G2	商品G1	商品G2	商品G1	商品G2	商品G1	商品G2	
...
i1	X	$C(i1,1)$	$C(i1,2)$	$C(i1,1)$	$C(i1,2)$	$C(i1,51)$	$C(i1,52)$

FIG. 5

【図 6】

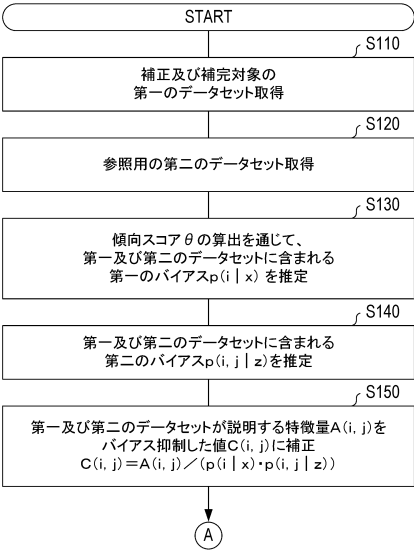


FIG. 6

【図 7】

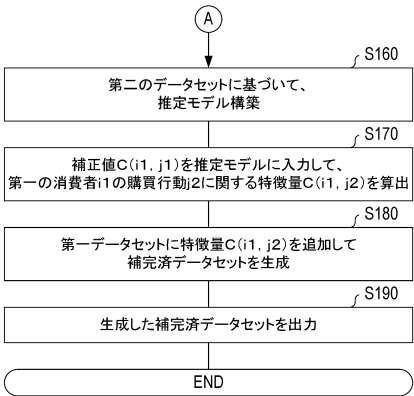


FIG. 7

10

20

30

40

50

フロントページの続き

東京都港区赤坂五丁目 3 番 1 号 株式会社博報堂 D Y ホールディングス内

審査官 上田 威

- (56)参考文献 特開 2 0 2 2 - 0 8 6 9 8 1 (J P , A)
米国特許出願公開第 2 0 1 7 / 0 0 1 7 9 7 5 (U S , A 1)
特開 2 0 1 6 - 0 6 2 5 1 1 (J P , A)
- (58)調査した分野 (Int.Cl. , D B 名)
G 0 6 Q 1 0 / 0 0 - 9 9 / 0 0